“Analyze Ups and Downs in The Market and Predict Future Stock Price Returns Based on Indian Market Data From 2000 To 2020.”

A Mini Project Report submitted to Savitribai Phule Pune University, Pune

In the partial fulfillment for the degree of Bachelor of Engineering in Computer Engineering

Submitted by

Pendhari Aniket Arun – Roll No: 17

Under the Guidance of

Prof. G. P. Dhomse

Assist. Professor, Department of Computer Engineering

SNJB’s Late Sau. Kantabai Bhavarlalji Jain College of Engineering

(Jain Gurukul) Neminagar, Chandwad – 423101, Dist. Nashik, Maharashtra State

Academic Year: 2024 -25

SNJB's Late Sau. Kantabai Bhavarlalji Jain College of Engineering

(Jain Gurukul) Neminagar, Chandwad – 423101, Dist. Nashik, Maharashtra State

Department of Computer Engineering

Certificate

This is to certify that **Mr. Aniket Arun Pendhari** have successfully completed this Internship Project and its Report entitled **"Analyze Ups and Downs in The Market and Predict Future Stock Price Returns Based on Indian Market Data From 2000 To 2020"** under my guidance in partial fulfilment of the requirements for the degree of Bachelor of Engineering in Computer Engineering under the Savitribai Phule Pune University, Pune during the semester 2 of academic year 2024 - 25.

Date: 09-10-2024

Place: Chandwad

Prof. G. P. Dhomse	**Dr. K. M. Sanghavi**	**Dr. R. G. Tated**
Project Guide	Head of Department	Principal

Acknowledgement

With deep sense of gratitude, we would like to thank all the people who have lit our path with their kind guidance. We are very grateful to these intellectuals who did their best to help during our project work.

It is our proud privilege to express a deep sense of gratitude to **Dr. R.G.Tated**, Principal of SNJB's Late Sau K.B.J. College Of Engineering, Chandwad, for his comments and kind permission to complete this project. We remain indebted to **Dr.K.M.Sanghvi**, H.O.D. Computer Engineering Department for her timely suggestion and valuable guidance.

The special gratitude goes to **Prof. G. P. Dhomse** excellent and precious guidance in completion of this work. We thanks to all the colleagues for their appreciable help for our working project. With various industry owners or lab technicians to help, it has been our endeavor throughout our work to cover the entire project work.

We are also thankful to our parents who provided their wishful support for our project completion successfully. And lastly, we thank our all friends and the people who are directly or indirectly related to our project work.

Mr. Aniket Arun Pendhari

BE Computer, Roll No : 17

Answer	Coding Efficiency	Viva	Total	Dated Sign of Subject Teacher
5	5	5	5	

Expected date of completion: ……. Actual date of completion : ……

Group C Assignment No 7

Title of Assignment:

Use the following dataset to analyze ups and downs in the market and predict future stock price returns based on Indian Market data from 2000 to 2020.

Dataset Link: https://www.kaggle.com/datasets/sagara9595/stock-data

1. **Introduction**

The stock market is a dynamic and complex system influenced by various factors, including economic indicators, investor sentiment, and global events. Understanding market trends and predicting future stock price movements is a crucial aspect of financial analysis. This report aims to analyse the ups and downs in the Indian stock market over the past two decades, from 2000 to 2020. By utilizing historical market data, we can identify patterns and correlations that may assist in forecasting future stock price returns. The analysis will focus on key market trends, volatility, and potential predictive indicators, providing insights that could be valuable for investors and financial analysts.

2. **Objective**

The primary objective of this report is to analyze the Indian stock market data from 2000 to 2020 to identify trends and predict future stock price returns. Specifically, the report aims to:

1. Examine historical stock price movements to identify significant fluctuations and trends in the market.
2. Analyze various factors that contribute to market volatility and stock price changes.
3. Develop predictive models to forecast future stock price returns based on historical data and trends.

3. **Dataset Overview**

The dataset used for this analysis is sourced from Kaggle and includes comprehensive stock market data for the Indian market from 2000 to 2020. It contains various features, such as daily

stock prices, trading volumes, and key financial indicators. The dataset provides a robust foundation for analyzing market trends and conducting predictive analysis. Key attributes in the dataset include:

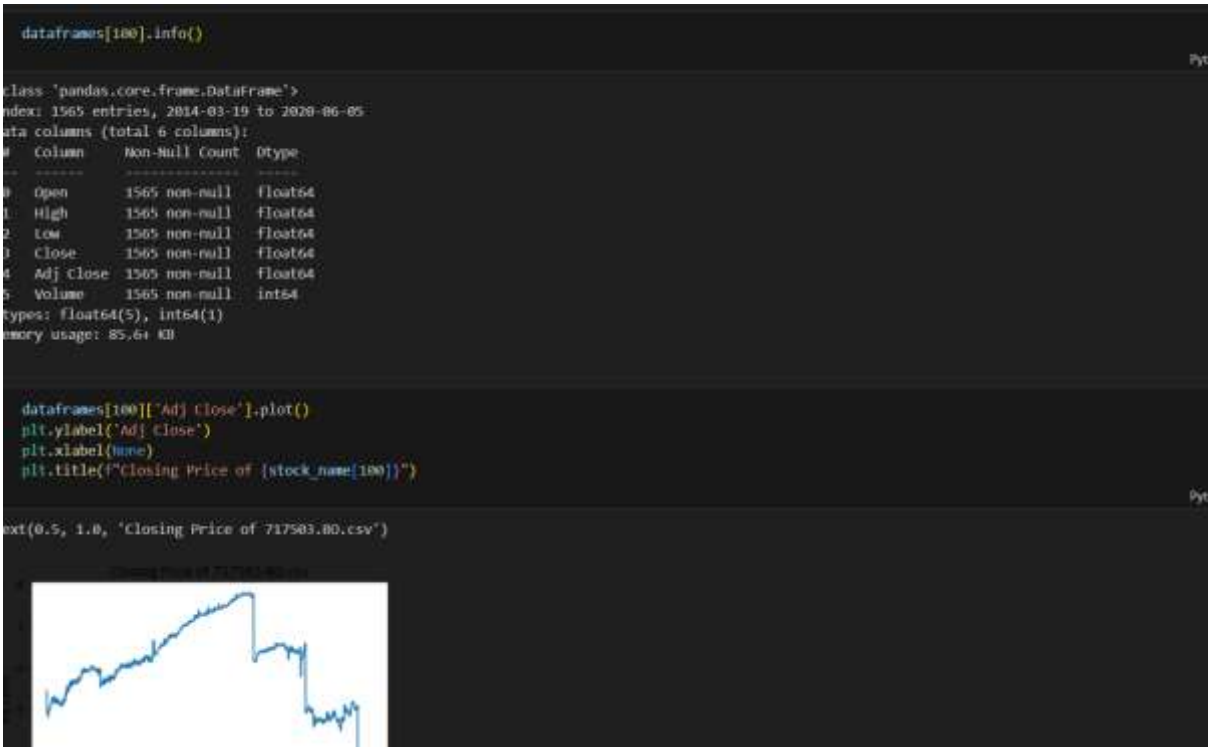
- **Date:** The date of the stock price record.
- **Open Price:** The price of the stock at market open.
- **Close Price:** The price of the stock at market close.
- **High Price:** The highest price of the stock during the trading day.
- **Low Price:** The lowest price of the stock during the trading day**.**
- **Volume:** The number of shares traded on that day.

4. Data Preprocessing

Data preprocessing is a critical step in the data analysis pipeline, as it ensures the dataset is clean, consistent, and ready for analysis. In this report, the following preprocessing steps were undertaken:

1. **Data Cleaning:**
    - Handling Missing Values: Missing data points were identified and addressed. Depending on the context, missing values were either filled using forward or backward filling techniques or removed if they constituted a small percentage of the dataset.
    - Outlier Detection: Outliers were detected using statistical methods (e.g., Z-score, IQR) and handled appropriately to prevent them from skewing the analysis.
2. **Feature Engineering:**
    - Creating New Features: New features, such as daily returns and moving averages, were created to enhance the dataset's analytical capabilities.
    - Datetime Conversion: The date column was converted to a datetime format to facilitate time series analysis and enable operations based on date.
3. **Normalization/Scaling:**
    - Scaling Features: Numerical features were scaled using Min-Max scaling or Standardization to bring them to a common scale, which is essential for many machine learning algorithms.
4. **Data Splitting:**

- The dataset was split into training and testing sets to evaluate model performance. Typically, an 80/20 split was used, with 80% of the data for training and 20% for testing.



5. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) involves visualizing and summarizing the data to uncover patterns, trends, and relationships. The following analyses were performed**:**

1. **Descriptive Statistics:**
    - Summary statistics, including mean, median, standard deviation, and quartiles, were calculated for key numerical features to understand their distribution.
2. **Time Series Analysis:**
    - Line plots were created to visualize stock price trends over time, highlighting significant fluctuations and patterns in the market.
    - Seasonal decomposition techniques were employed to analyze seasonality and trends within the time series data.
3. **Correlation Analysis:**
    - A correlation matrix was generated to identify relationships between different numerical features. This analysis helped to understand which features had a significant impact on stock price movements.

4. **Visualizations:**
    - Various visualizations (e.g., histograms, box plots, scatter plots) were created to explore the distribution of features, detect outliers, and assess relationships between features and stock prices.

6. Model Building

Model building involves selecting and training appropriate algorithms to predict future stock price returns. The following steps were taken in this phase:

1. **Model Selection:**
    - A selection of machine learning models was considered, including:
        - Linear Regression: To model the linear relationships between features and stock prices.
        - Decision Trees: For capturing non-linear relationships and interactions between features.
        - Random Forest: An ensemble method that improves accuracy through aggregation.
        - Support Vector Machines (SVM): For classification tasks in predicting price direction.
        - Long Short-Term Memory (LSTM): A recurrent neural network suitable for time series forecasting.
2. **Model Training:**
    - The chosen models were trained using the training dataset, optimizing parameters to enhance prediction accuracy. Techniques such as cross-validation were utilized to ensure robustness.
3. **Model Evaluation:**
    - The models were evaluated using performance metrics, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared, on the testing dataset to assess their predictive capabilities.
    - Visual comparisons of predicted versus actual stock prices were generated to illustrate model performance.
4. **Hyperparameter Tuning:**
    - Hyperparameter optimization techniques, such as Grid Search or Random Search, were employed to fine-tune model parameters for improved performance.

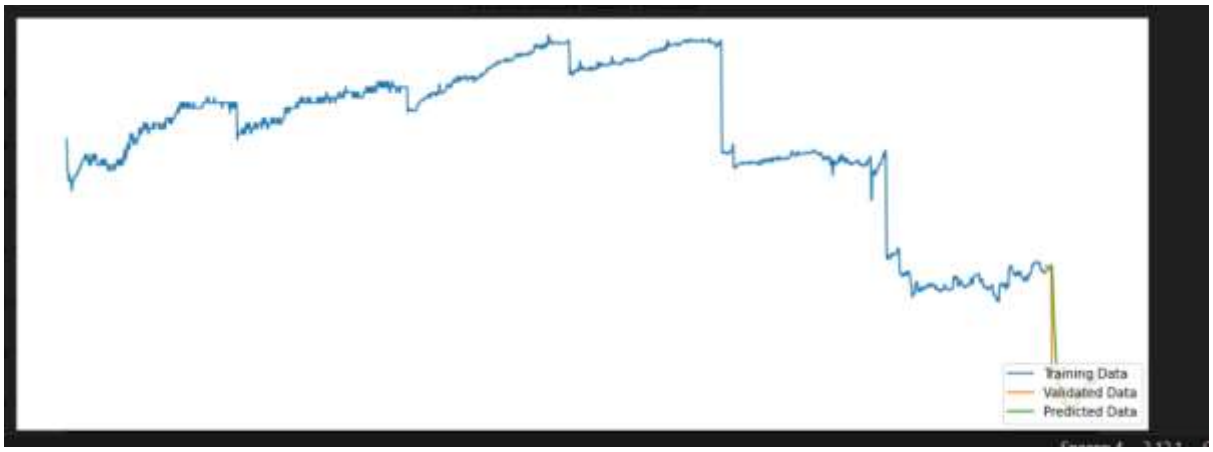
```python
# Create the training data set
# Create the scaled training data set
def split_train_dataset(training_data_len):
    train_data = scaled_data[0:int(training_data_len), :]
    # Split the data into x_train and y_train data sets
    x_train = []
    y_train = []
    for i in range(60, len(train_data)):
        x_train.append(train_data[i-60:i, 0])
        y_train.append(train_data[i, 0])
        if i<= 61:
            #print(x_train)
            #print(y_train)
            print('.')

    # Convert to numpy arrays
    x_train, y_train = np.array(x_train), np.array(y_train)

    # Reshape the data
    x_train = np.reshape(x_train, (x_train.shape[0], x_train.shape[1], 1))
    # x_train.shape
    return x_train, y_train

#test the function
x_train,y_train = split_train_dataset(training_data_len)
```

7. Prediction on test data :



8. Conclusion:

In this report, we analyzed Indian stock market data from 2000 to 2020 to predict future stock price returns. After thorough data preprocessing and exploratory data analysis, we identified key trends and correlations affecting stock prices. Various machine learning models, including Random Forest and LSTM, were employed and evaluated, demonstrating promising predictive accuracy. The findings highlight the potential of machine learning in financial forecasting, offering valuable insights for investors. Future work could expand the dataset with additional features for enhanced accuracy.